Jingwen Leng

Address
Computer Science and Engineering Department,
800 Dongchuan Road, Shanghai Jiao Tong University,
Shanghai, China, 200240

Contact Info.
Email: leng-jw@cs.sjtu.edu.cn
Mobile: 13262671153
Web: `www.jingwenleng.com`

Research Interests

Energy-efficiency and resilient architectures

- Heterogeneous computing:
  - efficiency and reliability optimization for GPU and accelerators
- Machine learning acceleration:
  - parallelization in distributed systems, and hardware acceleration

Academic Positions

Shanghai Jiao Tong University University, China **Dec 2016 - Present**
Assistant Professor, Computer Science and Engineering Department

Education

The University of Texas at Austin, Texas, USA **Dec 2012 - Dec 2016**
Ph.D., Electrical and Computer Engineering

- Advisor: Professor Vijay Janapa Reddi
- Area of Study: Computer Architecture, Compiler and VLSI

The University of Texas at Austin, Texas, USA **Aug 2010 - Dec 2012**
M.S., Electrical and Computer Engineering

- GPA: 3.8/4.0

Shanghai Jiao Tong University, Shanghai China **Sep 2006 - Jun 2010**
B.S., Electrical Engineering

- GPA 3.65/4.0

Awards and Honors

Best Paper Candidate in the International Symposium on Low Power Electronics and Design (ISLPED)	**2014**
Best of Computer Architecture Letters Award	**2014**
Excellent Undergraduate of Shanghai City	**2010**
Excellent Academic Awards of Shanghai Jiao Tong University	**2010, 2007**
First prize in China Undergraduate Mathematical Contest in Modeling	**2009**
National Scholarship (Top five percent out of over 600 students)	**2008**

Research Experience

The University of Texas at Austin, Texas, USA **Sept 2010 - Dec 2016**

Resilient GPU Architectures (continued in IBM Research)

- Developed *GPUVolt*, a validated GPU voltage simulator, to study voltage noise phenoma. It adopts a distributed grid for modeling the power delivery network.
- Analysis reveals two predominant types of voltage droops (local first-order and global second-order) in GPU architecture.
- Proposed a two-level voltage smoothing mechanism that can effectively smooth voltage droops.

Energy-effcient Big Data Engine (in progress)

- Building big data engine (BDE), a customized multi-GPU rack server for big data analysis to improve energy efficiency.

- Studying the energy-performance tradeoff for big data analytics workloads on GPUs.

Energy-efficient GPGPU Computing

- Studied the cause of energy-inefficiency in GPGPU programs, and exploited techniques at different granularities for energy optimization, such as the whole processor, groups of cores and SIMD lane level.
- Developed *GPUWattch*, a configurable, cycle-level and validated GPGPU power model using a circuit- to architecture-level bottom up methodology.
- Validated the power model against real hardware power measurements using custom-built power sensing circuitry.

Industry Experience

IBM T.J. Watson Research Center, New York, USA
Researh Intern, Manager: Pradip Bose **March 2015 - Dec 2016**

Cross-Layer Resilient GPU for Maximizing Energy Efficiency

- Perform V_{min} measurement on several off-the-shelf GPU cards to demonstrate that a large optimization opportunity exists at the voltage guardband level.
- Characterize the process, temperature and voltage variation impact on the V_{min}, and experimentally determine that the intra-kernel inductive voltage droop mostly determines the program's V_{min} value.
- Introduce the concept of *predictive guardbanding* that involves both software and hardware for guardband management, and demonstrate that the software can use performance counters to predict V_{min} accurately.

SRI International, Princeton, New Jersey, USA
Researh Intern, Vision Systems Group **Summer 2013**

- Studied and characterized the error-resilience nature of computer vision algorithms by developing an runtime error injecting framework (using Pin binary instrumentation compiler).
- Explored the trade-off among performance, energy efficiency and error resilience of vision algorithms.
- Proposed a lightweight check and recovery mechanism to improve the error-resilience of the algorithms.

Samsung Austin Research and Development Center, Austin, Texas, USA
Intern, GPU Logic Design Team **Summer 2011**

- Participated in the logic design and physical design of the ARM Mali-T604 GPU for the mobile market.
- Constructed a methodology to check clock gating efficiency based on gate level simulation.
- Developed an infrastructure to check the percentage of clock gates inserted by the synthesis tool in the post-layout gate list.

Publications

- ***Jingwen Leng***, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, Vijay Janapa Reddi, ***"Safe Limits on Voltage Reduction Efficiency in GPUs: a Direct Measurement Approach"***, in International Symposium on Microarchitecture (MICRO), Dec. 2015
- *Yazhou Zu, Charles Lefurgy,* ***Jingwen Leng***, Matthew Halpern, Michael Floyd, Vijay Janapa Reddi, ***"Adaptive Guardband Scheduling to Improve System-level Efficiency of POWER7+"***, in International Symposium on Microarchitecture (MICRO), Dec. 2015

- ***Jingwen Leng***, Yazhou Zu, Vijay Janapa Reddi, ***"GPU Voltage Noise: Characterizing and Smoothing Spatial and Temporal Voltage Noise Interference in GPU Architectures"***, to appear in International Symposium on High Performance Computer Architecture (HPCA), Feb. 2015
- ***Jingwen Leng***, Yazhou Zu, Minsoo Rhu, Meeta Gupta, Vijay Janapa Reddi, ***"GPUVolt: Characterizing and Mitigating Voltage Noise in GPUs"***, in International Symposium on Low Power Electronics and Design (ISLPED), 2014.
- ***Jingwen Leng***, Yazhou Zu, Vijay Janapa Reddi, ***"Energy Efficiency Benefits of Reducing the Voltage Guardband on the Kepler GPU Architecture"***, in 10th IEEE Workshop on Silicon Errors in Logic - System Effects (SELSE), 2014.
- Minsoo Rhu, Michael Sullivan, ***Jingwen Leng*** and Mattan Erez, ***"A Locality-Aware Memory Hierarchy for Energy-Efficient GPU Architectures"***, in International Symposium on Microarchitecture (MICRO), 2013.
- ***Jingwen Leng***, Syed Gilani, Ahmed El-Shafiey, Tayler Hetherington, Nam Sung Kim, Tor M. Aamodt, Vijay Janapa Reddi, ***"GPUWattch: Enabling Energy Optimizations in GPGPUs"***, in Proceedings of International Symposium on Computer Architecture (ISCA), 2013.
- Yuhao Zhu, Aditya Srikanth, ***Jingwen Leng***, Vijay Janapa Reddi, ***"Exploiting Webpage Characteristics for Energy-Efficient Mobile Web Browsing"***, Computer Architecture Letter (CAL), 2012.

Teaching Experience

Tutorials

- Jingwen Leng and Ahmed El-Shafiey, ***"GPUWattch + GPGPU-Sim: An Integrated Framework for Energy Optimizations in Manycore Architectures"***, International Symposium on Performance Analysis of Systems and Software (ISPASS), 2014.
- Jingwen Leng, Wilson Fung and Num Sung Kim, ***"GPUWattch + GPGPU-Sim: An Integrated Framework for Energy Optimizations in Manycore Architectures"***, International Symposium on Microarchitecture MICRO, 2013.
- Jingwen Leng and Tayler Hetherington, ***"GPUWattch + GPGPU-Sim: An Integrated Framework for Energy Optimizations in Manycore Architectures"***, International Symposium on Computer Architecture (ISCA), 2013.
- Jingwen Leng and Tayler Hetherington, ***"GPUWattch + GPGPU-Sim: An Integrated Framework for Energy Optimizations in Manycore Architectures"***, International Symposium on Performance Analysis of Systems and Software (ISPASS), 2013.

Classes

- **Teaching Assistant**, Code Generation and Optimization, The University of Texas at Austin, Fall 2013.
  - Designed programming assignments using LLVM compiler infrastructure for a class of 35 graduate students.
- **Teaching Assistant**, Advanced Embedded Micro-Controller System, The University of Texas at Austin, Fall 2011.

Tools

GPUWattch [http://gpuwattch.ece.utexas.edu]

- Architecture-level power model for GPUs.
- Open-sourced for academic research.
- Tutorials at major architecture conferences.

GPUVolt [http://gpuvolt.ece.utexas.edu]

- Distributed power delivery network simulator for GPUs.
- Open-sourced for academic research.

Skills

Languages: Proficient in C/C++, CUDA, Verilog, and Perl; Comfortable with JAVA, VHDL and Tcl.
Tools: GPGPU Simulator, McPAT power model, GNU development and debug tools, Cadence tools (Virtuoso, Encounter), Synopsys tools (Design Compiler, VCS, Prime Time), MATLAB, LabVIEW, LaTeX.

Professional Services

Reviewed papers for MICRO, HPCA, IPDPS, ISPASS.